

TP 0 — Partie 1: Découverte de l'interface d'Amazon Web Service (AWS)

TP 0 — Partie 1: Découverte de l'interface d'Amazon Web Service (AWS)

Objectifs

1. Création du compte AWS Academy
2. Exploration
- 3 Création d'un espace de stockage Amazon Simple Storage Service (S3)
4. Copie des données dans votre espace de stockage
5. Création d'une clef SSH
6. Création d'une machine virtuelle
7. Connexion à sa VM
8. Jouer avec sa VM
 - 8.1 Mise en place des fichiers du TP
 - 8.2 Benchmark des langages
 - 8.3 Un shell dans le navigateur
9. Eteindre sa VM

Objectifs

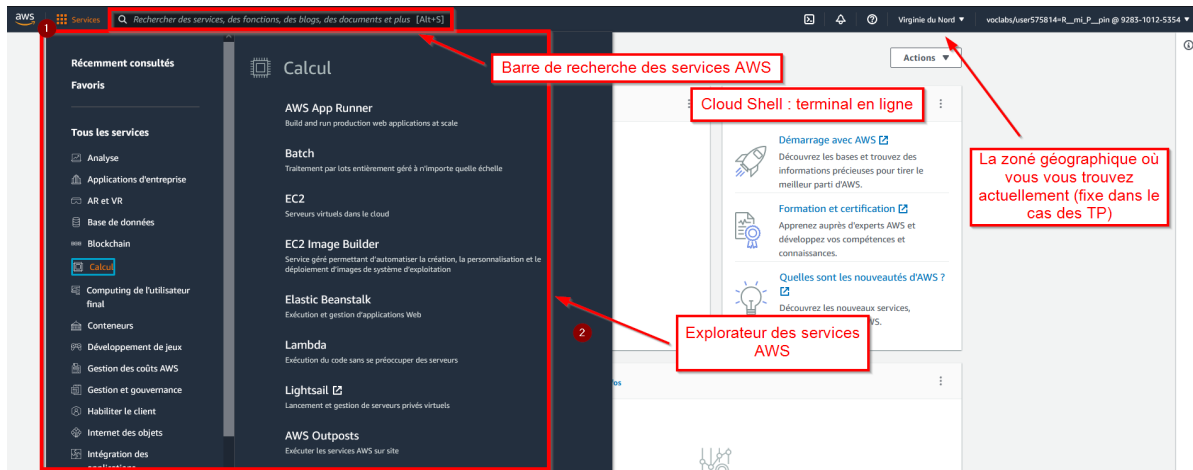
Ce TP a pour but de vous familiariser avec l'interface d'Amazon Web Service (AWS). Pendant ce TP vous allez :

- Créer un compte AWS educate si ce n'est pas déjà fait
- Copier des données dans votre espace de stockage Amazon Simple Storage Service (S3)
- Lancer une machine virtuelle (VM) et s'y connecter en SSH.
- Exécuter différentes commandes de base
 - `ls` pour lister les documents dans un dossier
 - `cd` pour change directory pour naviguer dans une arborescence de fichiers
 - `yum` pour installer un package
 - `aws s3 cp` pour copier des fichiers depuis S3
 - `chmod` pour changer les permissions d'un fichier
 - `time [commande]` pour mesurer la temps d'exécution d'une commande
- Eteindre votre VM

1. Création du compte AWS Academy

Suivez les instructions à partir du mail AWS Academy reçu sur votre adresse ENSAI pour créer votre compte si ce n'est pas fait, puis celles du fichier [AWS Academy Learner Lab - Student](#) Guide pour accéder à la console AWS.

Le compte AWS que vous allez utiliser pour les TP est localisé en Virginie du Nord. Ne changez pas cela ! Comme votre compte est à but purement scolaire, vous ne disposez pas de l'intégralité des services de la plateforme (vous n'avez pas accès aux information de facturation par exemple). De même votre compte est un compte généré par aws academy, et vous ne pouvez pas y accéder sans passez par ce service. Il est donc inutile de le noter. Pour des connexions futures à AWS dans le cadre scolaire, passez toujours par le portail AWS academy.



2. Exploration

Dans l'onglet "Services", trouvez, entre autres :

- EC2, les services de calcul
- S3, les services de calcul
- La section dédiée aux bases de données
- La section dédiée au *machine-learning*
- La section dédiée à l'analyse de données

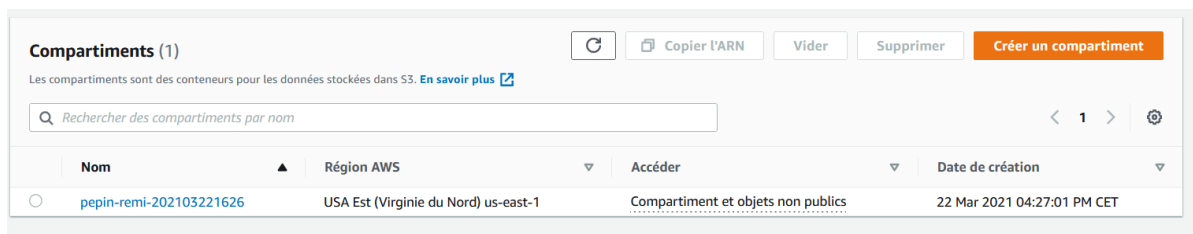
3 Création d'un espace de stockage Amazon Simple Storage Service (S3)

Amazon Simple Storage Service (S3) est la solution de base que propose AWS pour stocker vos données de manière pérenne. Amazon dit assurer une durabilité de vos données de 99,999999999 %. Cela signifie que si vous stockez 10 000 000 objets avec Amazon S3, vous pouvez vous attendre à perdre en moyenne un objet unique une fois tous les 10 000 ans.

Ce stockage est assuré à coût relativement élevé (de l'ordre de ~0,02 \$/Go/mois), sachant que vous payez en sus les opérations de lecture (de l'ordre de ~0,09 \$/Go ; les écritures sont gratuites). 1 To de données vous coûte ainsi 240€ à l'année. Pour comparaison, un disque dur externe SSD d'1 To coûte actuellement ~100€ (pour une durabilité moindre), et un cloud-storage (type dropbox) pour particulier coûte ~10€ / mois pour 2 To (pour une durabilité comparable). S3 est ainsi destiné à des données utilisées régulièrement par d'autres application hébergées par AWS. D'autres offres de stockage existent comme les archives, pour des données utilisées moins régulièrement, ou les bases de données.

Tous les services que vous propose AWS peuvent nativement lire depuis et écrire vers S3 si vous leur en donnez le droit. Ainsi, les programmes que vous exécutez et les données que vous traitez peuvent être importés/exportés dans S3. Chaque élément hébergé dans S3, appelé "objet", est accessible par une URL **unique**. Vous pouvez restreindre ou au contraire étendre les droits d'accès à vos objets.

☐ Dans la barre de recherche, cherchez "S3" et cliquez dessus



- ☐ Cliquez sur "Créer un compartiment" (en anglais un "bucket")
- ☐ Choisissez un nom unique à votre compartiment (comme votre nom-prénom et la date et heure du jour)
- ☐ Laissez toutes les valeurs par défaut et descendez en bas de la page pour créer votre compartiment

4. Copie des données dans votre espace de stockage

- ☐ Cliquez sur le nom de votre compartiment pour aller sur sa page

- ☐ À partir du bouton "Charger", ajoutez tous le fichier pour le lab0 disponible sur Moodle.

- ☐ Une fois le chargement terminé cliquez sur votre fichier. Vous arriverez sur une page similaire avec les lien pour accéder à votre fichier. Aussi bien S3 que HTTP

Présentation de l'objet	
Propriétaire remi.pepin	URI S3 s3://pepin-remi-202103221626/rule20210219-184802.jsonl
Région AWS USA Est (Virginie du Nord) us-east-1	Nom de la ressource Amazon (ARN) arn:aws:s3::pepin-remi-202103221626/rule20210219-184802.jsonl
Dernière modification 22 Mar 2021 04:46:05 PM CET	Balise d'entité (Etag) 433dfec8145842d4a565ec9033be30fa
Taille 904,0 o	URL de l'objet https://pepin-remi-202103221626.s3.amazonaws.com/rule20210219-184802.jsonl
Type jsonl	
Clé rule20210219-184802.jsonl	

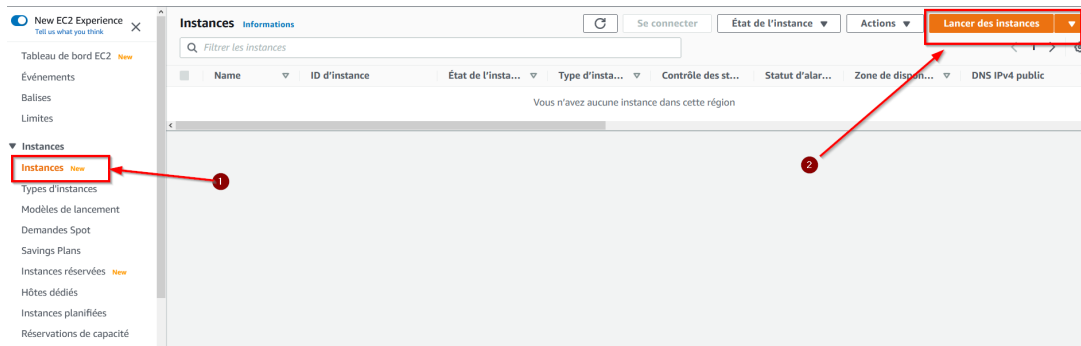
5. Création d'une clef SSH

SSH (Secure SHell) permet de se connecter de façon sécurisée à un système Unix comme le cluster de l'Ensaï qui est sous une distribution Linux. Pour plus d'information, je vous conseille de lire le début de cette [page web](#)

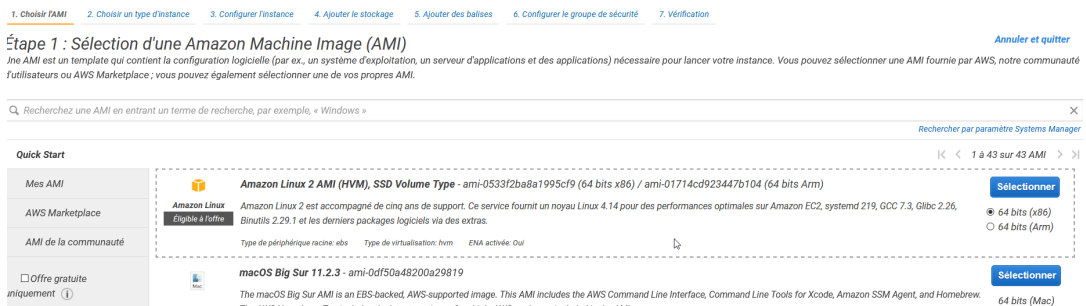
- ☐ Dans la barre de recherche, cherchez "paire de clés" et cliquez dessus
- ☐ Cliquez sur "Créer une paire de clés"
- ☐ Donnez lui un nom (par ex: "ensaï_big_data_TP"), sélectionnez le format PPK si vous utilisez une machine windows, et pem si vous utilisez une machine sous Linux / Mac Os, et cliquez sur "créer"
- ☐ Cela va lancer le téléchargement d'un fichier, ne le perdez pas !

6. Création d'une machine virtuelle

- ☐ Dans la barre de recherche, cherchez "EC2" et cliquez dessus



- ☐ Cliquez sur **Lancer une instance**
- ☐ Vous devez choisir l'image de la machine à créer appelé ici AMI pour *Amazon Machine Image*. Une image contient notamment le système d'exploitation. Choisissez la première: **Amazon Linux 2 AMI (HVM) - Kernel 5.10**.



- ☐ Vous choisissez ensuite la configuration de VM. Par exemple, vous pouvez choisir une machine d'usage général à 1 cœurs **t2.micro** pour avoir une machine de faible puissance mais peu chère (0.012\$/heure) pour une machine plus puissante comme une **t2.xlarge** (0.188\$/heure). Comme la facturation est au temps d'utilisation, pensez à

éteindre vos machines à la fin du TP !

Étape 2 : Choisir un type d'instance

Actuellement sélectionné : t2.micro (- ECU, 1 vCPU, 2.5 GHz, 1 Go mémoire, EBS uniquement)

	Famille	Type	vCPU	Mémoire (Go)	Stockage d'instance (Go)	Disponible en version optimisée pour EBS	Performances réseau	Prise en charge IP
<input type="checkbox"/>	t2	t2.nano	1	0.5	EBS uniquement	-	Faibles à modérées	Oui
<input checked="" type="checkbox"/>	t2	t2.micro <small>Éligible à l'offre gratuite</small>	1	1	EBS uniquement	-	Faibles à modérées	Oui
<input type="checkbox"/>	t2	t2.small	1	2	EBS uniquement	-	Faibles à modérées	Oui
<input type="checkbox"/>	t2	t2.medium	2	4	EBS uniquement	-	Faibles à modérées	Oui
<input type="checkbox"/>	t2	t2.large	2	8	EBS uniquement	-	Faibles à modérées	Oui
<input type="checkbox"/>	t2	t2.xlarge	4	16	EBS uniquement	-	Modérées	Oui
<input type="checkbox"/>	t2	t2.2xlarge	8	32	EBS uniquement	-	Modérées	Oui
<input type="checkbox"/>	t3	t3.nano	2	0.5	EBS uniquement	Oui	Jusqu'à 5 gigabits	Oui
<input type="checkbox"/>	t3	t3.micro	2	1	EBS uniquement	Oui	Jusqu'à 5 gigabits	Oui
<input type="checkbox"/>	t3	t3.small	2	2	EBS uniquement	Oui	Jusqu'à 5 gigabits	Oui

[Annuler](#) [Précédent](#) [Vérifier et lancer](#) [Suivant : Configurer les détails de l'instance](#)

- ☐ Sur l'écran suivant sélectionnez pour le rôle IAM `LabInstanceProfile`

Number of instances ⓘ [Launch into Auto Scaling Group](#) ⓘ

Purchasing option ⓘ ☐ Request Spot instances

Network ⓘ [Create new VPC](#)

Subnet ⓘ [Create new subnet](#)

Auto-assign Public IP ⓘ

Hostname type ⓘ

DNS Hostname ⓘ ☒ Enable IP name IPv4 (A record) DNS requests
☒ Enable resource-based IPv4 (A record) DNS requests
☐ Enable resource-based IPv6 (AAAA record) DNS requests

Placement group ⓘ ☐ Add instance to placement group

Capacity Reservation ⓘ

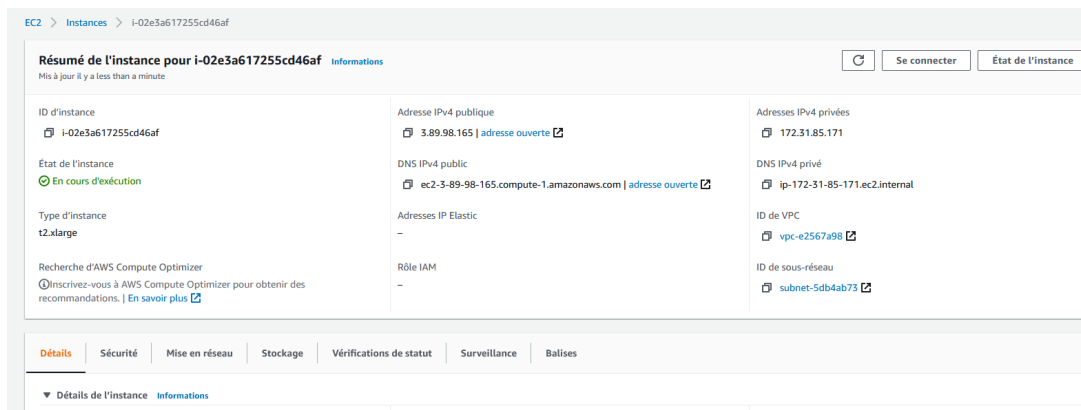
Domain join directory ⓘ [Create new directory](#)

IAM role ⓘ [Create new IAM role](#)

- ☐ Puis cliquez sur `Vérifier et lancer`
- ☐ Ensuite validez la création de l'instance.
- ☐ Choisissez la bonne paire de clés
- ☐ Et voilà! Votre VM est en cours de lancement. Cliquez sur `Affichez les instances` et attendez quelques minutes !

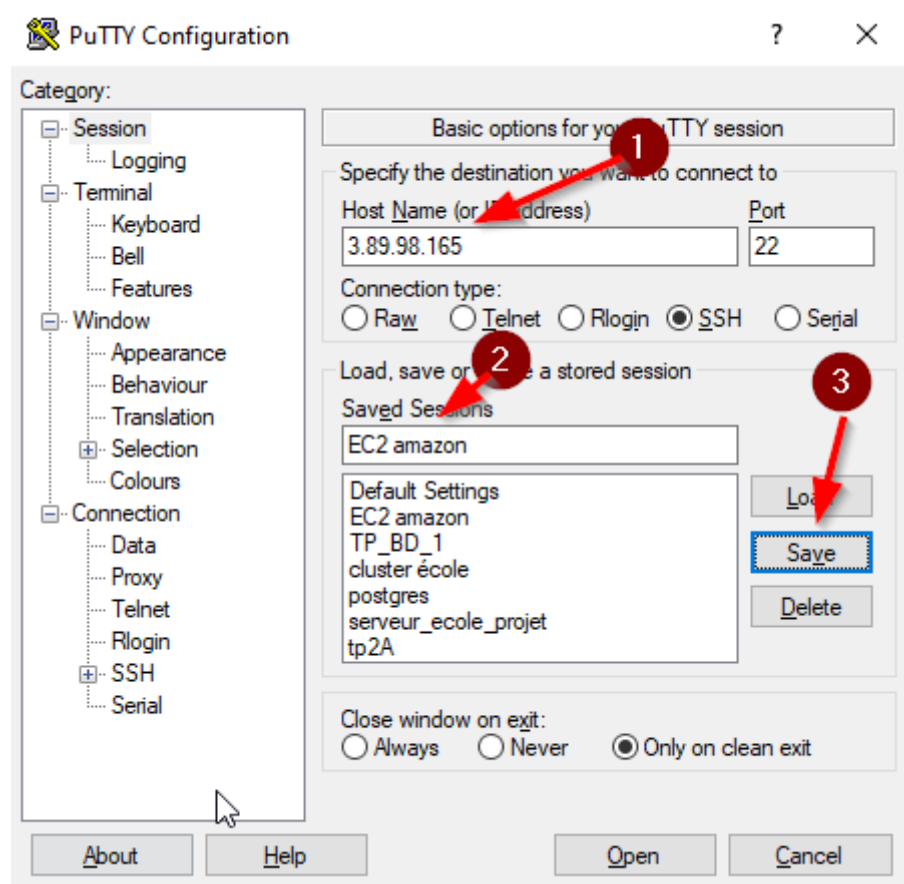
7. Connexion à sa VM

- ☐ Une fois l'instance lancée vous pouvez accéder à son écran d'administration en cliquant sur son `id d'instance`

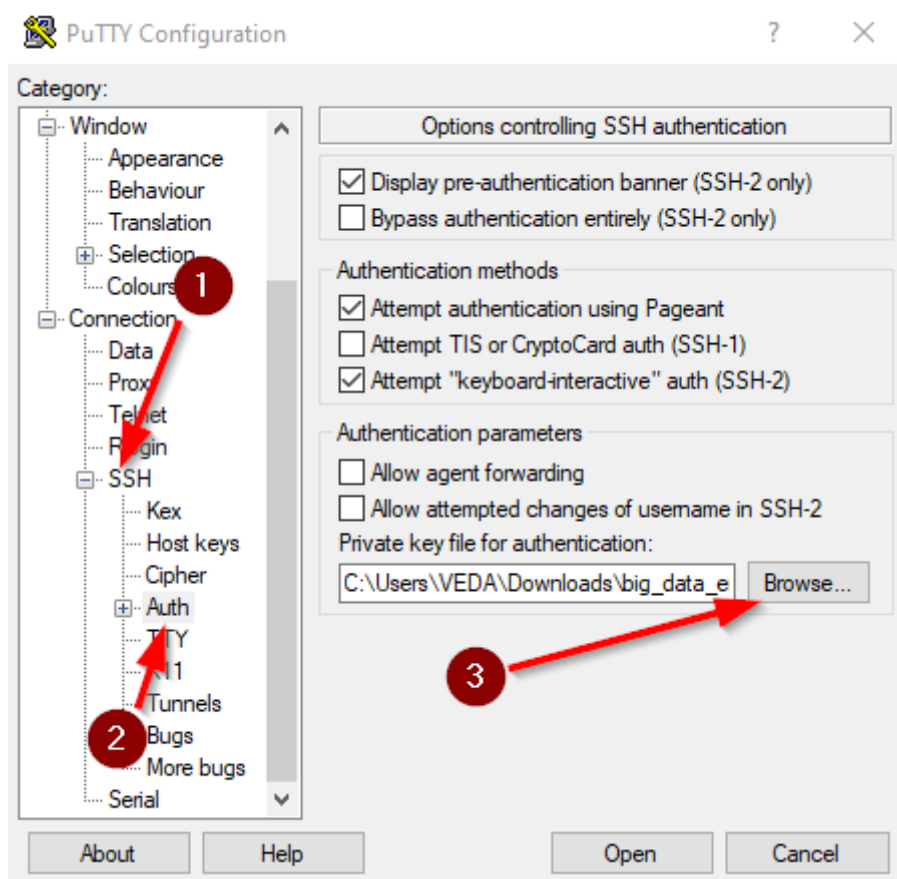


Vous y trouverez de nombreuses informations, mais surtout l'adresse **IPv4 publique** qui est l'adresse IP (Internet Protocol) de votre machine pour y accéder en étant à l'extérieur de la plateforme AWS, par exemple depuis votre ordinateur.

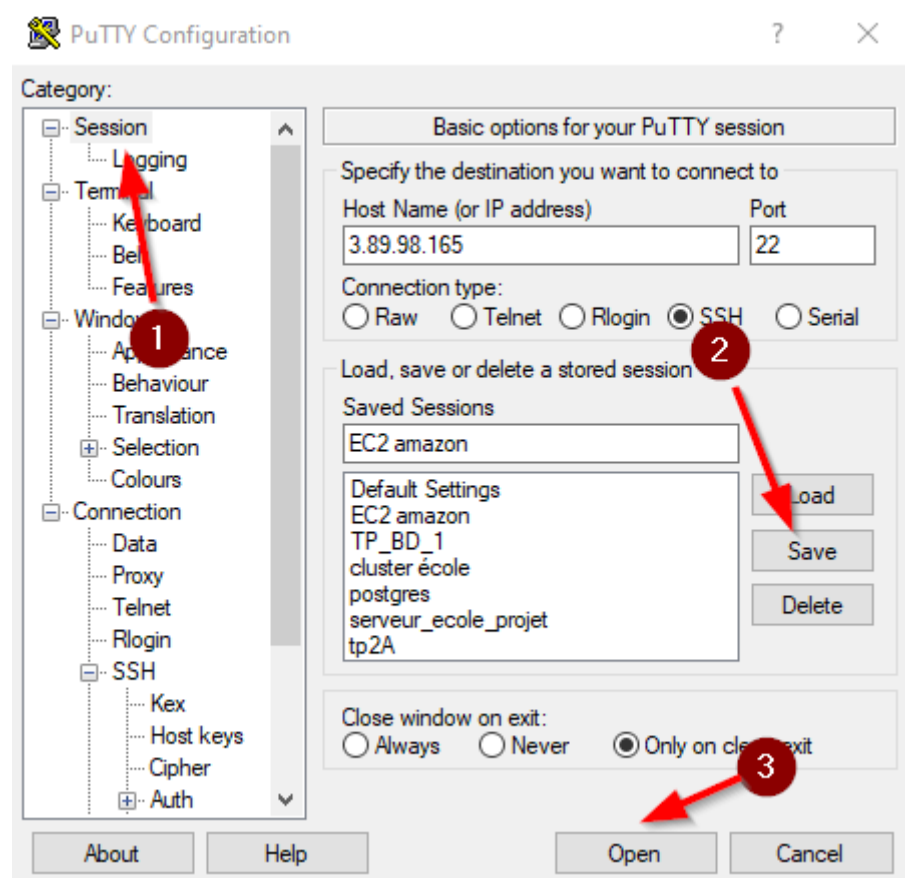
- ☐ Lancez PuTTY
- ☐ Dans la partie **Host Name** saisissez l'adresse publique de votre serveur
- ☐ Dans la partie **saved session** rentrez le nom que vous voulez puis cliquez sur **Save**



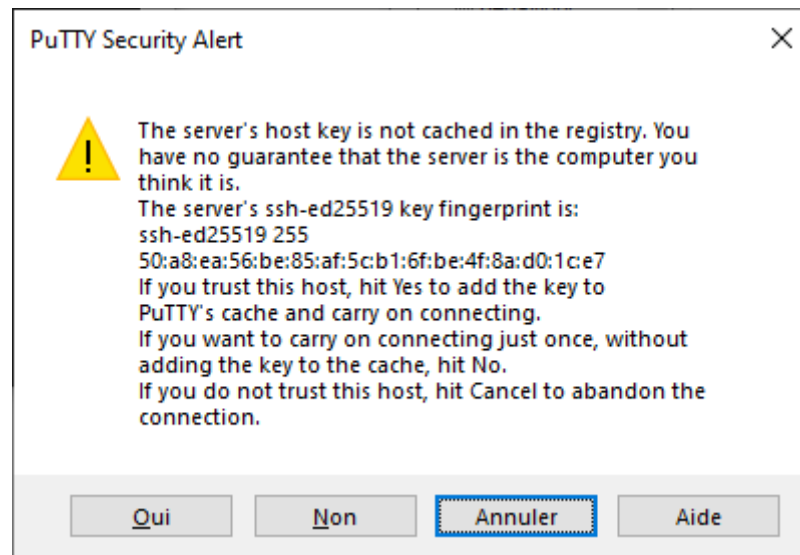
- ☐ Puis allez dans le menu **SSH** et chargez votre fichier .ppk



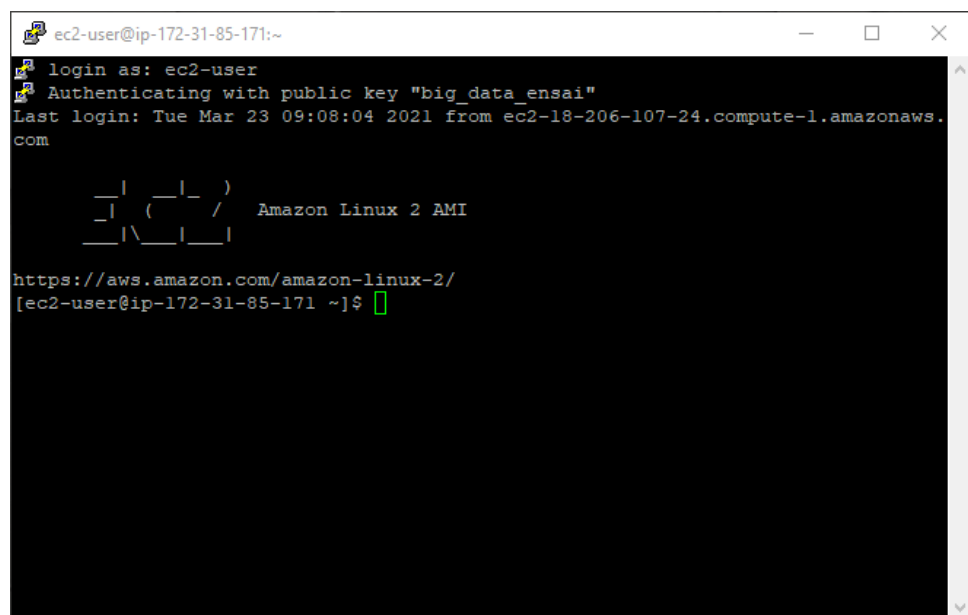
- ☐ Enfin retournez dans l'écran initial, cliquez sur le nom de la session puis sur **save** pour sauvegarder votre configuration. Lancez la session SSH.



- ☐ Une fenêtre semblable va s'ouvrir pour vous prévenir que c'est la première fois que vous vous connectez à cette machine et si vous lui faites confiance. Vous allez cliquer sur **oui**



- ☐ Un terminal va s'ouvrir avec écrit `login as: .` Saisissez `ec2-user` puis validez ([documentation officielle](#))



- ☐ Voilà vous venez de vous connecter à votre machine virtuelle. **Bien que visuellement le terminal se trouve sur votre écran, tout ce que vous allez exécuter dans se terminal sera réalisé sur une machine distante.** Vous pouvez ainsi réaliser de calculs très longs et nécessitant une grande puissance de calcul sur une machine puissance depuis votre ordinateur. Par contre cette machine n'a pas d'interface graphique (GUI : *graphical user interface*) et va nécessiter de connaître quelques rudiments de *bash*.

8. Jouer avec sa VM

Le but de cette section es de vous faire manipuler quelques commandes de base en *bash*. C'est sûrement nouveau pour vous, et peut-être effrayant, mais il est important que vous ayant au moins des notions de base.

8.1 Mise en place des fichiers du TP

- ☐ Téléchargez vos fichiers stockés sur S3. Pour ce faire vous allez saisir la commande suivante `aws s3 cp [s3://object/uri] [output/fo1der]`. Pour récupérer l'URI de votre objet S3, allez sur la page de votre objet et cliquez sur "Copier l'URI S3". Pour `output/fo1der`, vous allez utiliser le répertoire courant avec un `.`. Vous devrait obtenir

une commande et un retour similaire à celui ci

```
1 [ec2-user@ip-172-31-85-99 ~]$ aws s3 cp s3://remi-pepin-21032022/fichiersTP0.zip .
2 download: s3://remi-pepin-21032022/fichier TP.zip to ./fichier TP.zip
```

- ☐ Avec la commande `ls` (*list*) vérifiez que vous avez bien téléchargé les fichiers sur S3 dans le répertoire courant.
- ☐ Vous allez maintenant extraire les fichiers de l'archive avec la commande `unzip [nom de votre fichier]`. Vérifiez que cela à bien fonctionné avec la commande `ls`
- ☐ Maintenant que vous avez vos fichiers, vous allez exécuter le script `get_data.sh`. Pour ce faire tapez `./get_data.sh`. Vous devriez obtenir le résultat suivant

```
-bash: ./get_data.sh: Permission denied
```

Pour des raisons de sécurité, ce fichier ne peut être exécuté pour le moment. Vous pouvez vérifier cela en faisant la commande `ls -l get_data.sh` qui devrait produire ceci

```
-rw-rw-r-- 1 ec2-user ec2-user 459 Mar 21 09:28 get_data.sh
```

Le premier `rw-` signifie que le propriétaire (vous) du fichier à des droit de lecture (r) et d'écriture (w) mais pas d'exécution. Le second dit la même chose pour le groupe, et le dernier `r--` signifie les autres utilisateurs ont juste des droits de lecture.

Vous allez remédier à cela avec la commande `chmod` (*change mode*) qui permet de changer les permissions sur fichier. Et vous allez seulement vous (*user*) donnez les droit d'exécution sur les fichiers `get_data.sh`, `awk.sh` et `GetMaxTempC` avec la commande `chmod 764 get_data.sh awk.sh GetMaxTempC`. Si vous refaites la commande `ls -l get_data.sh` vous allez obtenir

```
-rwxrw-r-- 1 ec2-user ec2-user 459 Mar 21 09:28 get_data.sh
```

Le x signifiant droit d'exécution. Vous pouvez désormais exécuter le fichier `get_data.sh` qui va vous récupérer les fichiers nécessaire au benchmark.

8.2 Benchmark des langages

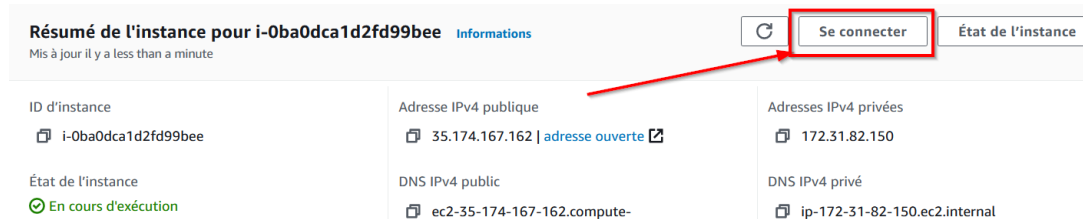
Dans cette partie vous allez reproduire l'expérience du cours consistant à tester la vitesse de traitement de différents langages. Cela va se faire essentiellement avec la commande `time`. La commande `time` permet de mesurer la temps d'exécution d'une commande passer en argument. Exemple `time chmod 764 get_data.sh` permet de mesurez le temps nécessaire pour pour changer les permission du fichier `get_data.sh`. Notez chacun des résultats et vérifiez qu'ils sont cohérents avec ceux du cours. Si ce n'est pas les cas, essayez de comprendre pourquoi.

- ☐ Pour lancer le code C compilé et le script bash `awk.sh` vous avez juste à les appeler en prefixant leur nom par `./` pour dire que vous souhaitez appeler le fichier et pas une commande.
- ☐ Pour lancer un jar vous devez utiliser la commande `java -jar [file.jar]`
- ☐ Pour les codes python utilisez la commande `python3`
- ☐ Pour pouvoir utiliser Cython, il faut :

- ☐ Installer `python-devel` pour pouvoir créer des extensions python. Pour ce faire, vous allez utiliser `yum`, un gestionnaire de package pour certaines distributions linux. La commande à réaliser est `sudo yum install -y python3-devel.x86_64` (`sudo` pour dire que vous exécutez la commande en super user, `yum` pour dire que vous utilisez le gestionnaire de package, `install` pour dire que vous voulez installer un package, `-y` pour valider l'installation, et `python3-devel.x86_64` le nom du package)
- ☐ Installer `cython` avec `pip3` et compiler le code cython en faisant :
 - ☐ `cd cythond_code` pour *change directory* qui permet de se déplacer dans votre arborescence
 - ☐ `python3 setup.py` pour lancer la compilation
 - ☐ `cd ../` pour retourner dans le dossier parent.
- ☐ Vous pouvez maintenant mesurer le temps que prend le code Cython en l'appelant avec `time` et `python3`
- ☐ R n'est pas par défaut installé. Pour l'installer vous allez utiliser le gestionnaire de package d'Amazon `amazon-linux-extras`, avec la ligne de commande suivante : `sudo amazon-linux-extras install R4 -y`. Le terminal va se remplir de texte pendant quelques minutes n'y prêtez pas attention, c'est juste la machine qui vous dit ce qu'elle fait. Ensuite pour lancer un script R vous devez seulement saisir `Rscript [filename.R]` dans votre terminal.

8.3 Un shell dans le navigateur

- ☐ Fermez votre terminal et retournez sur la page de votre instance EC2. Nous allons maintenant nous y connecter via un *cloud shell*. Pour ce faire cliquez sur `Se connecter`



Résumé de l'instance pour i-0ba0dca1d2fd99bee Informations

Mis à jour il y a less than a minute

ID d'instance i-0ba0dca1d2fd99bee	Adresse IPv4 publique 35.174.167.162 adresse ouverte	Adresses IPv4 privées 172.31.82.150
État de l'instance En cours d'exécution	DNS IPv4 public ec2-35-174-167-162.compute-	DNS IPv4 privé ip-172-31-82-150.ec2.internal

Vous allez arriver sur une page similaire à celle ci-dessous. Cliquez sur `Se connecter`

EC2 > Instances > i-0ba0dca1d2fd99bee > Se connecter à l'instance

Se connecter à l'instance [Informations](#)

Se connecter à votre instance à i-0ba0dca1d2fd99bee l'aide de l'une de ces options

Connexion d'instance EC2

Session Manager

Client SSH

ID d'instance
i-0ba0dca1d2fd99bee

Adresse IP publique
35.174.167.162

Nom utilisateur

Connectez-vous à l'aide d'un nom d'utilisateur personnalisé ou du nom d'utilisateur par défaut ec2-user de l'AMI utilisée pour lancer l'instance.

Remarque : Dans la plupart des cas, le nom d'utilisateur deviné est correct. Cependant, lisez les instructions d'utilisation de l'AMI pour vérifier si le propriétaire de l'AMI a modifié le nom d'utilisateur par défaut.

Annuler

Se connecter

Après quelques instants vous allez arriver sur votre *cloud shell*

```
□ Last login: Tue Mar 23 11:17:26 2021 from ec2-18-206-107-24.compute-1.amazonaws.com

  _ | _ | _ )
 _ | ( _ | /  Amazon Linux 2 AMI
 _ | \ _ | _ |

https://aws.amazon.com/amazon-linux-2/
[ec2-user@ip-172-31-82-150 ~]$ █

I

i-0ba0dca1d2fd99bee
Public IPs: 35.174.167.162 Private IPs: 172.31.82.150
```

Depuis cette écran vous êtes connecté à votre machine distante. Par exemple tapez la commande suivante `ls` pour voir que vous avez bien vos fichiers, puis tentez de les réexécuter.

9. Eteindre sa VM

Le coût d'une VM est fonction de son temps d'utilisation, pas du travail qu'il accomplit. Ainsi, une fois le travail effectué, vous *devez* éteindre vos VMs ! **Même si le coût horaire est bas, faire tourner une machine EC2 pendant 1 semaine se chiffre en dizaines d'euros!**

Pour éteindre votre VM, allez sur la page d'accueil `EC2 > Instances` en cours d'exécution ou sur la bar de navigation `Instances > Instances`, enfin `Etat de l'instance`. Selon le type d'instance, vous pouvez l'arrêter (**EN: stop**, pour la réutiliser plus tard), ou la résilier (**EN: terminate**, i.e. la supprimer). Dans les deux cas, les données en mémoire et le stockage local sont perdus, mais dans le premier cas, la configuration (URL et IP) sont conservés.

New EC2 Experience
Tell us what you think

Tableau de bord EC2 New

Événements

Balises

Limites

▼ Instances

Instances New

Types d'instances

Modèles de lancement

Demandes Spot

Savings Plans

Arrêt réussi i-0ba0dca1d2fd99bee

Instances (1/3) Informations

Se connecter

État de l'instance ▲

Actions ▼

Lancer

Filter les instances

	Name	ID d'instance	État de l'insta...	Type d'insta
<input type="checkbox"/>	-	i-02e3a617255cd46af	Arrêté(e)	t2.large
<input type="checkbox"/>	-	i-048005c7bcd927adc	Arrêté(e)	t2.micro
<input checked="" type="checkbox"/>	-	i-0ba0dca1d2fd99bee	En cours d'ex...	t2.micro

Arrêter l'instance

Démarrer l'instance

Redémarrer l'instance

Mettre l'instance en veille prolongée

Résilier l'instance

4/4 vérifications n... Aucune at...

1

2

3